

# Multiprocessor Approximate Message Passing with Column-Wise Partitioning\*

Yanting Ma  
North Carolina State University  
yma7@ncsu.edu

Yue M. Lu  
Harvard University  
yuelu@seas.harvard.edu

Dror Baron  
North Carolina State University  
barondror@ncsu.edu

## Abstract

Solving a large-scale regularized linear inverse problem using multiple processors is important in various real-world applications due to the limitations of individual processors and constraints on data sharing policies. This paper focuses on the setting where the matrix is partitioned column-wise. We extend the algorithmic framework and the theoretical analysis of approximate message passing (AMP), an iterative algorithm for solving linear inverse problems, whose asymptotic dynamics are characterized by state evolution (SE). In particular, we show that column-wise multiprocessor AMP (C-MP-AMP) obeys an SE under the same assumptions when the SE for AMP holds. The SE results imply that (i) the SE of C-MP-AMP converges to a state that is no worse than that of AMP and (ii) the asymptotic dynamics of C-MP-AMP and AMP can be identical. Moreover, for a setting that is not covered by SE, numerical results show that damping can improve the convergence performance of C-MP-AMP.

## 1 Introduction

Many scientific and engineering problems can be modeled as solving a regularized linear inverse problem of the form

$$y = Ax + w, \quad (1)$$

where the goal is to estimate the unknown  $x \in \mathbb{R}^N$  given the matrix  $A \in \mathbb{R}^{n \times N}$  and statistical information about the signal  $x$  and the noise  $w \in \mathbb{R}^n$ .

In some scenarios, it might be desirable to partition the matrix  $A$  either column-wise or row-wise and store the sub-matrices at different processors. The partitioning style depends on data availability, computational considerations, and privacy concerns. For example, in high-dimensional settings where  $N \gg n$ , or in situations where the columns of  $A$ , which represent features in feature selection problems [2], cannot be shared among processors for privacy preservation, column-wise partitioning might be preferable. In this paper, we consider multiprocessor computing for the (non-overlapping) column-wise partitioned linear inverse problem:

$$y = \sum_{p=1}^P A_p x_p + w, \quad (2)$$

where  $P$  is the number of processors,  $A_p \in \mathbb{R}^{n \times N_p}$  is the sub-matrix that is stored in Processor  $p$ , and  $\sum_{p=1}^P N_p = N$ .

Many studies on solving the column-wise partitioned linear inverse problem (2) have been in the context of distributed feature selection. Zhou *et al.* [3] modeled feature selection as a parallel group

---

\*This document serves as a supporting document for [1]

testing problem. Wang *et al.* [4] proposed to de-correlate the data matrix before partitioning, and each processor then works independently using the de-correlated matrix without communication with other processors. Peng *et al.* [5] studied problem (2) in the context of optimization, where they proposed a greedy coordinate-block descent algorithm and a parallel implementation of the fast iterative shrinkage-thresholding algorithm (FISTA) [6].

Our work is based on the approximate message passing (AMP) framework [7]. AMP is an efficient iterative algorithm for solving linear inverse problems (1). In the large scale random setting, its average asymptotic dynamics are characterized by a state evolution (SE) formalism [8], which allows one to accurately predict the average estimation error at every iteration. Recently, a finite-sample analysis of AMP [9] showed that when the prior distribution of the input signal  $x$  has i.i.d. sub-Gaussian entries,<sup>1</sup> the average performance of AMP concentrates to the SE prediction at an exponential rate in the signal dimension  $N$ .

Our goal is to extend the AMP algorithmic framework and the SE analysis in [9] to the column-wise partitioned linear inverse problem (2). We show that column-wise multiprocessor AMP (C-MP-AMP) obeys a new SE under the same model assumptions where the SE for AMP holds. With the new SE, we can predict the average estimation error in each processor at every iteration. Moreover, the comparison between the SE of AMP and that of C-MP-AMP implies that (i) the estimation error of C-MP-AMP is no worse than that of AMP and (ii) with a specific communication schedule between the processors and the fusion center that coordinates the processors, the asymptotic dynamics of C-MP-AMP are identical to that of AMP. This result implies a speedup linear in the number of processors.

It is worth mentioning that row-wise multiprocessor AMP [10–12] obeys the same SE as AMP, because it distributes the computation of matrix-vector multiplication among multiple processors and aggregates the results before any other operations. Some existing work on row-wise multiprocessor AMP [12–14] introduces lossy compression to the communication between processors and the fusion center, whereas we assume perfect communication and focus on the theoretical justifications and implications of the new SE of C-MP-AMP.

The remainder of the paper is organized as follows. Section 2 introduces the C-MP-AMP algorithm (Algorithm 1), the state evolution sequences, and our main performance guarantee (Theorem 1), which is a concentration result for PL loss functions acting on the outputs generated by Algorithm 1 concentrates to the state evolution prediction. Section 3 proves Theorem 1. The proof is mainly based on Lemmas 3 and 4. The proof of Lemma 3 is the same as in [9] using the result that we prove in Lemma 2. Section 4 proves Lemma 4.

## 2 Column-Wise Multiprocessor AMP and State Evolution

### 2.1 Review of AMP

Approximate message passing (AMP) [7] is a fast iterative algorithm for solving linear inverse problems (1). Starting with an all-zero vector  $x^0$  as its initial estimate, at the  $t$ th iteration, AMP proceeds according to

$$z^t = y - Ax^t + \frac{z^{t-1}}{n} \sum_{i=1}^N \eta'_{t-1}([x^{t-1} + A^* z^{t-1}]_i), \quad (3)$$

$$x^{t+1} = \eta_t(x^t + A^* z^t), \quad (4)$$

---

<sup>1</sup>A random variable  $X$  is sub-Gaussian if there exist positive constants  $c$  and  $\kappa$  such that  $P(|X - \mathbb{E}X| > \epsilon) \leq ce^{-\kappa\epsilon^2}, \forall \epsilon > 0$ .

where vectors with negative iteration indices are all-zero vectors,  $A^*$  denotes the transpose of a matrix  $A$ ,  $\eta_t : \mathbb{R} \rightarrow \mathbb{R}$  is a Lipschitz function with weak derivative  $\eta'_t$ , for any  $u \in \mathbb{R}^N$ ,  $[u]_i$  denotes its  $i$ th entry. The function  $\eta_t$  acts coordinate-wise when applied to vectors. That is, the vector  $(\eta_t(u_1), \eta_t(u_2), \dots, \eta_t(u_N))$  is denoted by  $\eta_t(u)$ .

Under the assumptions on the measurement matrix  $A$ , the signal  $x$ , the measurement noise  $w$ , and the denoising function  $\eta_t(\cdot)$  as listed in [9, Section 1.1], the sequence of the estimates  $\{x^t\}$  that is generated by AMP (3) (4) has the following property [9]. For all  $\epsilon \in (0, 1)$ , there exist constants  $K_t, \kappa_t > 0$  independent of  $n$  or  $\epsilon$ , such that

$$P \left( \left| \frac{1}{N} \sum_{i=1}^N \phi(x_i^{t+1}, x_i) - \mathbb{E} [\phi(\eta_t(X + \tau^t Z), X)] \right| \geq \epsilon \right) \leq K_t e^{-\kappa_t n \epsilon^2}, \quad (5)$$

where  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a pseudo-Lipschitz function of order 2 (PL(2)),<sup>2</sup>  $X \sim p_X$ ,  $Z$  is a standard normal random variable that is independent of  $X$ , and  $\tau^t$  is defined via the following recursion  $((\sigma^0)^2 = \delta^{-1} \mathbb{E}[X^2], \delta = n/N)$ :

$$\begin{aligned} (\tau^t)^2 &= \sigma_W^2 + (\sigma^t)^2, \\ (\sigma^{t+1})^2 &= \delta^{-1} \mathbb{E} \left[ (\eta_t(X + \tau^t Z) - X)^2 \right]. \end{aligned} \quad (6)$$

Notice that (5) implies, by applying the Borel-Cantelli Lemma, the almost sure convergence result proved in [8]:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \phi(x_i^{t+1}, x_i) \stackrel{a.s.}{=} \mathbb{E} [\phi(\eta_t(X + \tau^t Z), X)]. \quad (7)$$

If we choose  $\phi(x, y) = (x - y)^2$ , then (7) characterizes the mean square error (MSE) achieved by AMP at each iteration.

## 2.2 Column-Wise Multiprocessor AMP

In our proposed column-wise multiprocessor AMP (C-MP-AMP) algorithm, the fusion center collects vectors that represent the estimations of the portion of the measurement vector  $y$  contributed by the data from individual processors according to a pre-defined communication schedule. The sum of these vectors is computed in the fusion center and transmitted to all processors. Each processor performs standard AMP iterations with a new equivalent measurement vector, which is computed using the vector received from the fusion center. The pseudocode for C-MP-AMP is presented in Algorithm 1.

## 2.3 Performance Guarantee

Similar to AMP, the dynamics of the C-MP-AMP algorithm can be characterized by an SE formula. Let  $(\sigma_p^{0, \hat{k}_0})^2 = \delta_p^{-1} \mathbb{E}[X^2]$ , where  $\delta_p = n/N_p$ ,  $\forall p = 1, \dots, P$ . For outer iterations  $1 \leq s \leq \hat{s}$  and inner

---

<sup>2</sup>Recall the definition of PL(2) from [8]: a function  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  is said to be PL(2) if there is  $L > 0$  such that  $|f(x) - f(y)| \leq L(1 + \|x\| + \|y\|)\|x - y\|$ ,  $\forall x, y \in \mathbb{R}^m$ , where  $\|\cdot\|$  denotes the Euclidean norm.

---

**Algorithm 1** C-MP-AMP

---

**Inputs to Processor  $p$ :**  $y, A_p, \{\hat{k}_s\}_{s=0,\dots,\hat{s}}$  (maximum number of inner iterations at each outer iteration).

**Initialization:**  $x_p^{0,\hat{k}_0} = 0, z_p^{0,\hat{k}_0-1} = 0, r_p^{0,\hat{k}_0} = 0, \forall p$ .

**for**  $s = 1 : \hat{s}$  **do** (loop over outer iterations)

At the fusion center:  $g^s = \sum_{u=1}^P r_u^{s-1,\hat{k}_{s-1}}$

At Processor  $p$ :

$x_p^{s,0} = x_p^{s-1,\hat{k}_{s-1}}, r_p^{s,0} = r_p^{s-1,\hat{k}_{s-1}}$

**for**  $t = 0 : \hat{t}_s - 1$  **do** (loop over inner iterations)

$z_p^{s,k} = y - g_s - (r_p^{s,k} - r_p^{s,0})$

$x_p^{s,k+1} = \eta_{s,k}(x_p^{s,k} + A_p^* z_p^{s,k})$

$r_p^{s,k+1} = A_p x_p^{s,k+1} - \frac{z_p^{s,k}}{n} \sum_{i=1}^{N_p} \eta'_{s,k}([x_p^{s,k} + A_p^* z_p^{s,k}]_i).$

**Output from Processor  $p$ :**  $x_p^{\hat{s},\hat{k}_{\hat{s}}}.$

---

iterations  $0 \leq t \leq \hat{k}_s$ , we define the sequences  $\{(\sigma_p^{s,k})^2\}$  and  $\{(\tau_p^{s,k})^2\}$  as

$$(\sigma_p^{s,0})^2 = (\sigma_p^{s-1,\hat{k}_s})^2, \quad (8)$$

$$(\tau_p^{s,k})^2 = \sigma_W^2 + \sum_{u \neq p} (\sigma_u^{s,0})^2 + (\sigma_p^{s,k})^2, \quad (9)$$

$$(\sigma_p^{s,k+1})^2 = \delta_p^{-1} \mathbb{E} \left[ \left( \eta_{s,k}(X + \tau_p^{s,k} Z) - X \right)^2 \right], \quad (10)$$

where  $Z$  is a standard normal random variable that is independent of  $X$ .

With these definitions, we have the following performance guarantee for C-MP-AMP.

**Theorem 1.** *Under the assumptions listed in [9, Section 1.1], let  $P$  be a fixed integer, for  $p = 1, \dots, P$ , let  $n/N_p = \delta_p \in (0, \infty)$  be a constant. Define  $N = \sum_{p=1}^P N_p$ . Then for any  $PL(2)$  function  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ , we have  $\forall \epsilon \in (0, 1)$ , there exist constants  $K_{s,k}, \kappa_{s,k} > 0$  independent of  $n, \epsilon$ , such that*

$$P \left( \left| \frac{1}{N_p} \sum_{i=1}^{N_p} \phi(x_{p,i}^{s,k+1}, x_{p,i}) - \mathbb{E} \left[ \phi(\eta_{s,k}(X + \tau_p^{s,k} Z), X) \right] \right| \geq \epsilon \right) \leq K_{s,k} e^{-\kappa_{s,k} n \epsilon^2}, \forall p,$$

where  $x_p^{s,k+1}$  is generated by Algorithm 1,  $\tau_p^{s,k}$  is defined in (8–10),  $X \sim p_X$ , and  $Z$  is a standard normal random variable that is independent of  $X$ .

*Remark 1:* C-MP-AMP converges to a fixed point that is no worse than that of AMP. This statement can be demonstrated as follows. When C-MP-AMP converges, the quantities in (8–10) do not keep changing, hence we can drop all the iteration indices for fixed point analysis. Notice that the last term on the right hand side (RHS) of (9) vanishes, which leaves the RHS independent of  $p$ . That is,  $(\tau_p^{s,k})^2$  are equal for all  $p$ , hence we can further drop the processor index for  $(\tau_p^{s,k})^2$ . Denote  $(\tau_p^{s,k})^2$  by  $\tau^2$  for all  $s, k, p$ , and plug (10) into (9), then

$$\begin{aligned} \tau^2 &= \sigma_W^2 + \sum_{p=1}^P \delta_p^{-1} \mathbb{E} \left[ (\eta(X + \tau Z) - X)^2 \right] \\ &\stackrel{(a)}{=} \sigma_W^2 + \delta^{-1} \mathbb{E} \left[ (\eta(X + \tau Z) - X)^2 \right], \end{aligned}$$

which is identical to the fixed point equation obtained from (6). In the above, step (a) holds because  $\sum_{p=1}^P \delta_p^{-1} = \sum_{p=1}^P \frac{N_p}{n} = \frac{N}{n}$ . Because AMP always converges to the worst fixed point of the fixed point equation (6) [15], the average asymptotic performance of C-MP-AMP is identical to AMP when there is only one solution to the fixed point equation, and at least as good as AMP in case of multiple fixed points.

*Remark 2: The asymptotic dynamics of C-MP-AMP can be identical to AMP with a specific communication schedule.* This can be achieved by letting  $\hat{k}_s = 1, \forall s$ . In this case, the quantity  $(\tau_p^{s,k})$  is involved only for  $t = 0$ . Because the last term in (9) is 0 when  $t = 0$ , the computation of  $(\tau_p^{s,0})^2$  is independent of  $p$ . Therefore,  $\tau_p^{s,0}$  are again equal for all  $p$ . Dropping the processor index for  $(\tau_p^{s,k})^2$ , the recursion in (8–10) can be simplified as

$$\begin{aligned} (\tau^{s,0})^2 &= \sigma_W^2 + \sum_{p=1}^P \delta_p^{-1} \mathbb{E} \left[ (\eta_{s,0}(X + \tau^{s,0}Z) - X)^2 \right] \\ &= \sigma_W^2 + \delta^{-1} \mathbb{E} \left[ (\eta_{s-1,0}(X + \tau^{s-1,0}Z) - X)^2 \right], \end{aligned}$$

where the iteration evolves over  $s$ , which is identical to (6) evolving over  $t$ .

*Remark 3: Theorem 1 implies almost sure convergence.* Similar to the performance guarantee for AMP [9], the concentration result in Theorem 1 implies

$$\lim_{N \rightarrow \infty} \frac{1}{N_p} \sum_{i=1}^{N_p} \phi(x_{p,i}^{s,k+1}, x_{p,i}) \stackrel{a.s.}{=} \mathbb{E} \left[ \phi(\eta_{s,k}(X + \tau_p^{s,k}Z), X) \right], \forall p,$$

by the Borel-Cantelli Lemma.

### 3 Proofs of Theorem 1

Our proof follows closely from the proof for AMP in [9], with additional dependence structure to be addressed due to vectors being transmitted among processors.

#### 3.1 Proof Notations

Without loss of generality, we assume the sequence  $\{\hat{k}_s\}_{s \geq 0}$  in Algorithm 1 to be a constant value  $\hat{k}$ . Let  $t = s\hat{k} + k$ ,  $\theta(t) = \lfloor t/\hat{k} \rfloor \hat{k}$ . Given  $w \in \mathbb{R}^n$ ,  $x_p \in \mathbb{R}^{N_p}$ , for  $p = 1, \dots, P$ , define the column vectors  $h_p^{t+1}, q_p^t \in \mathbb{R}^{N_p}$  and  $b_p^t, m_p^t \in \mathbb{R}^n$  for  $t \geq 0$  recursively as follows. Starting with initial condition  $q_p^0 \in \mathbb{R}^{N_p}$ :

$$\begin{aligned} h_p^{t+1} &= A_p^* m_p^t - q_p^t, & q_p^t &= f_t(h_p^t, x_p) \\ b_p^t &= A_p q_p^t - \lambda_p^t m_p^{t-1}, & m_p^t &= b_p^t + \sum_{u \neq p} b_u^{\theta(t)} - w \end{aligned} \quad (11)$$

where

$$f_t(h^t, x_p) = \eta_{t-1}(x_p - h^t) - x_p, \quad \text{and} \quad \lambda_p^t := \frac{1}{\delta_p N_p} \sum_{i=1}^{N_p} f'_t(h_{p,i}^t, x_{p,i}). \quad (12)$$

In (12), the derivative of  $f_t : \mathbb{R}^2 \rightarrow \mathbb{R}$  is with respect to the first argument. We assume that  $\eta_t$  is Lipschitz for all  $t \geq 0$ , then it follows that  $f_t$  is Lipschitz for all  $t \geq 0$ . Consequently, the weak derivative and  $f'_t$  exist. Further,  $f'_t$  is assumed to be differentiable, except possibly at a finite number

of points, with bounded derivative whenever it exists. In (11), quantities with negative indices or with index  $\theta(t) = 0$  (i.e.,  $t < \hat{k}$ ) are defined to be zeros.

To see the equivalence between Algorithm 1 and the recursion defined in (11) and (12), we let  $x_p^0 = 0$ ,  $r_p^0 = 0$ ,  $z_p^t = 0$ , and

$$\begin{aligned} h_p^{t+1} &= x_p - (A^* z_p^t + x_p^t), & q_p^t &= x_p^t - x_p, \\ b_p^t &= r_p^t - A_p x_p, & m_p^t &= -z_p^t. \end{aligned}$$

Let  $(\sigma_p^0)^2 = \delta_p^{-1} \mathbb{E}[X^2]$ . We assume that  $(\sigma_p^0)^2$  is strictly positive for all  $p = 1, \dots, P$  and for all  $\epsilon \in (0, 1)$ , there exist  $K, \kappa > 0$  such that

$$P \left( \left| \frac{\|q_p^0\|^2}{n} - (\sigma_p^0)^2 \right| \geq \epsilon \right) \leq K e^{-\kappa n \epsilon^2}, \quad \forall p = 1, \dots, P. \quad (13)$$

Define the state evolution scalars  $\{\tau_p^t\}_{t \geq 0}$  and  $\{\sigma_p^t\}_{t \geq 1}$  for the the recursion defined in (11) as follows:

$$(\tau_p^t)^2 = (\sigma_p^t)^2 + \sum_{u \neq p} (\sigma_u^{\theta(t)})^2 + \sigma_W^2, \quad (\sigma_p^t)^2 = \frac{1}{\delta_p} \mathbb{E} \left[ (f_t(\tau_p^{t-1} Z, X))^2 \right], \quad (14)$$

where  $Z \sim \mathcal{N}(0, 1)$  and  $X \sim p_X$  are independent. Notice that with the equivalence between Algorithm 1 and the recursion 11, the state evolution scalars defined in (14) matches (8) - (10).

Writing the updating equations for  $b_p^t, h_p^{t+1}$  defined in (11) in matrix form, we have

$$X_p^t = A_p^* M_p^t, \quad Y_p^t = A_p Q_p^t, \quad (15)$$

where

$$\begin{aligned} X_p^t &= [h_p^1 + q_p^0 | h_p^2 + q_p^1 | \dots | h_p^t + q_p^{t-1}], & Y_p^t &= [b_p^0 | b_p^1 + \lambda_p^1 m_p^0 | \dots | b_p^{t-1} + \lambda_p^{t-1} m_p^{t-2}] \\ M_p^t &= [m_p^0 | m_p^1 | \dots | m_p^{t-1}], & Q_p^t &= [q_p^0 | q_p^1 | \dots | q_p^{t-1}]. \end{aligned}$$

Let  $(m_p^t)_{||}$  and  $(q_p^t)_{||}$  denote the projection of  $m_p^t$  and  $q_p^t$  onto the column space of  $M_p^t$  and  $Q_p^t$ , respectively. That is,

$$\begin{aligned} (m_p^t)_{||} &= M_p^t ((M_p^t)^* M_p^t)^{-1} (M_p^t)^* m_p^t \\ (q_p^t)_{||} &= Q_p^t ((Q_p^t)^* Q_p^t)^{-1} (Q_p^t)^* q_p^t. \end{aligned}$$

Let

$$\alpha_p^t = (\alpha_{p,0}^t, \alpha_{p,1}^t, \dots, \alpha_{p,t-1}^t)^*, \quad \gamma_p^t = (\gamma_{p,0}^t, \gamma_{p,1}^t, \dots, \gamma_{p,t-1}^t)^* \quad (16)$$

be the coefficient vectors of these projections. That is,

$$\alpha_p^t = ((M_p^t)^* M_p^t)^{-1} (M_p^t)^* m_p^t, \quad \gamma_p^t = ((Q_p^t)^* Q_p^t)^{-1} (Q_p^t)^* q_p^t. \quad (17)$$

and

$$(m_p^t)_{||} = \sum_{i=0}^{t-1} \alpha_{p,i}^t m_p^i, \quad (q_p^t)_{||} = \sum_{i=0}^{t-1} \gamma_{p,i}^t q_p^i. \quad (18)$$

Define

$$(m_p^t)_{\perp} = m_p^t - (m_p^t)_{||}, \quad (q_p^t)_{\perp} = q_p^t - (q_p^t)_{||}. \quad (19)$$

The main lemma will show that  $\alpha_p^t$  and  $\gamma_p^t$  concentrate around some constant  $\hat{\alpha}_p^t$  and  $\hat{\gamma}_p^t$ , respectively. We define these constants in the following subsection.

### 3.2 Concentrating Constants

Let  $\{\tilde{Z}_p^t\}_{t \geq 0}$  and  $\{\check{Z}_p^t\}_{t \geq 0}$  each be a sequence of zero-mean jointly Gaussian random variables whose covariance is defined recursively as follows. For  $t, r \geq 0$ ,

$$\mathbb{E}[\tilde{Z}_p^r \tilde{Z}_p^t] = \frac{\tilde{E}_p^{r,t}}{\sigma_p^r \sigma_p^t}, \quad \mathbb{E}[\check{Z}_p^r \check{Z}_p^t] = \frac{\check{E}_p^{r,t}}{\tau_p^r \tau_p^t}, \quad (20)$$

where

$$\begin{aligned} \tilde{E}_p^{r,t} &= \tilde{E}_p^{r,t} + \sum_{u \neq p} \tilde{E}_u^{\theta(r), \theta(t)} + \sigma_W^2, \\ \tilde{E}_p^{r,t} &= \delta_p^{-1} \mathbb{E} \left[ f_r(\tau_p^{r-1} \tilde{Z}_p^{r-1}, X) f_t(\tau_p^{t-1} \tilde{Z}_p^{t-1}, X) \right]. \end{aligned} \quad (21)$$

Moreover,  $\tilde{Z}_p^r$  is independent of  $\tilde{Z}_q^t$  and  $\check{Z}_p^r$  is independent of  $\check{Z}_q^t$  for all  $r, t \geq 0$  whenever  $p \neq q$ . Note that according to the definition of  $\sigma_p^t$  and  $\tau_p^t$  in (14), we have  $\tilde{E}_p^{t,t} = (\tau_p^t)^2$ ,  $\check{E}_p^{t,t} = (\sigma_p^t)^2$ , and  $\mathbb{E}[(\tilde{Z}_p^t)^2] = \mathbb{E}[(\check{Z}_p^t)^2] = 1$ . In (21), quantities with negative indices or with either  $\theta(t) = 0$  or  $\theta(r) = 0$  are zeros.

Define matrices  $\tilde{C}_p^t, \check{C}_p^t \in \mathbb{R}^{t \times t}$ ,  $p = 1, 2$ , such that

$$[\tilde{C}_p^t]_{r+1, s+1} = \tilde{E}_p^{r,s}, \quad [\check{C}_p^t]_{r+1, s+1} = \check{E}_p^{r,s}, \quad \forall r, s = 0, \dots, t-1.$$

Define vectors  $\tilde{E}_p^t, \check{E}_p^t \in \mathbb{R}^t$ ,  $p = 1, 2$ , such that

$$\tilde{E}_p^t = (\tilde{E}_p^{0,t}, \tilde{E}_p^{1,t}, \dots, \tilde{E}_p^{t-1,t}), \quad \check{E}_p^t = (\check{E}_p^{0,t}, \check{E}_p^{1,t}, \dots, \check{E}_p^{t-1,t}).$$

Define the concentrating values  $\hat{\alpha}_p^t$  and  $\hat{\gamma}_p^t$  as

$$\hat{\gamma}_p^t = (\tilde{C}_p^t)^{-1} \tilde{E}_p^t, \quad \hat{\alpha}_p^t = (\check{C}_p^t)^{-1} \check{E}_p^t. \quad (22)$$

Let  $(\sigma_p^0)_\perp^2 = (\sigma_p^0)^2$  and  $(\tau_p^0)_\perp^2 = (\tau_p^0)^2$ , and for  $t > 0$ , define

$$\begin{aligned} (\sigma_p^t)_\perp^2 &= (\sigma_p^t)^2 - (\hat{\gamma}_p^t)^* \tilde{E}_p^t = (\sigma_p^t)^2 - (\tilde{E}_p^t)^* (\tilde{C}_p^t)^{-1} \tilde{E}_p^t, \\ (\tau_p^t)_\perp^2 &= (\tau_p^t)^2 - (\hat{\alpha}_p^t)^* \check{E}_p^t = (\tau_p^t)^2 - (\check{E}_p^t)^* (\check{C}_p^t)^{-1} \check{E}_p^t. \end{aligned} \quad (23)$$

**Lemma 1.** *The matrices  $\tilde{C}_p^t$  and  $\check{C}_p^t$ ,  $\forall t \geq 0$ , defined above are invertible, and the scalars  $(\sigma_p^t)_\perp^2$  and  $(\tau_p^t)_\perp^2$ ,  $\forall t \geq 0$ , defined above are strictly positive.*

*Proof.* The proof for  $\tilde{C}_p^t$  being invertible and  $(\sigma_p^t)_\perp^2$  being strictly positive is the same as in [9]. Now consider  $\check{C}_p^{t+1}$ . Notice that  $\check{C}_p^{t+1}$  is the sum of a positive definite matrix  $(\check{C}_p^{t+1})$  and  $P$  positive semi-definite matrices, hence,  $\check{C}_p^{t+1}$  is positive definite. Consequently,

$$\det(\check{C}_p^{t+1}) = \det(\check{C}_p^t) \det((\tau_p^t)^2 - (\check{E}_p^t)^* (\check{C}_p^t)^{-1} \check{E}_p^t) > 0, \quad (24)$$

which implies  $(\tau_p^t)^2 - (\check{E}_p^t)^* (\check{C}_p^t)^{-1} \check{E}_p^t = (\tau_p^t)_\perp^2 > 0$ .  $\square$

### 3.3 Condition Distribution Lemma

Let the sigma algebra  $\mathcal{S}^{t_1, t}$  be generated by  $x, w, b_p^0, \dots, b_p^{t_1-1}, m_p^0, \dots, m_p^{t_1-1}, h_p^1, \dots, h_p^t, q_p^0, \dots, q_p^t, \forall p$ . We now compute the conditional distribution of  $A_p$  given  $\mathcal{S}^{t_1, t}$  for  $1 \leq p \leq P$ , where  $t_1$  is either  $t$  or  $t+1$ .

Notice that conditioning on  $\mathcal{S}^{t_1, t}$  is equivalent to conditioning on the linear constraints:

$$A_p Q_p^{t_1} = Y_p^{t_1}, \quad A_p^* M_p^t = X_p^t, \quad 1 \leq p \leq P, \quad (25)$$

where in (25), only  $A_p$ ,  $1 \leq p \leq P$ , are treated as random.

Let  $\mathbf{P}_{Q_p^{t_1}}^\parallel = Q_p^{t_1}((Q_p^{t_1})^* Q_p^{t_1})^{-1} Q_p^{t_1}$  and  $\mathbf{P}_{M_p^t}^\parallel = M_p^t((M_p^t)^* M_p^t)^{-1} M_p^t$ , which are the projectors onto the column space of  $Q_p^{t_1}$  and  $M_p^t$ , respectively. The following lemma provides the conditional distribution of the matrices  $A_p$ ,  $p = 1, \dots, P$ , given  $\mathcal{G}^{t_1, t}$ .

**Lemma 2.** *For  $t_1 = t$  or  $t+1$ , the conditional distribution of the random matrices  $A_p$ ,  $p = 1, \dots, P$ , given  $\mathcal{S}^{t_1, t}$  satisfies*

$$(A_1, \dots, A_P)|_{\mathcal{S}^{t_1, t}} \stackrel{d}{=} (\mathbf{E}_1^{t_1, t} + \mathbf{P}_{M_1^t}^\perp \tilde{A}_1 \mathbf{P}_{Q_1^{t_1}}^\perp, \dots, \mathbf{E}_P^{t_1, t} + \mathbf{P}_{M_P^t}^\perp \tilde{A}_P \mathbf{P}_{Q_P^{t_1}}^\perp)$$

where  $\mathbf{P}_{Q_p^{t_1}}^\perp = \mathbf{I} - \mathbf{P}_{Q_p^{t_1}}^\parallel$  and  $\mathbf{P}_{M_p^t}^\perp = \mathbf{I} - \mathbf{P}_{M_p^t}^\parallel$ .  $\tilde{A}_p \stackrel{d}{=} A_p$  and  $\tilde{A}_p$  is independent of  $\mathcal{S}^{t_1, t}$ . Moreover,  $\tilde{A}_p$  is independent of  $\tilde{A}_q$  for  $p \neq q$ .  $E_p^{t_1, t}$  is defined as

$$\begin{aligned} \mathbf{E}_p^{t_1, t} &= Y_p^{t_1}((Q_p^{t_1})^* Q_p^{t_1})^{-1} (Q_p^{t_1})^* + M_p^t((M_p^t)^* M_p^t)^{-1} (X_p^t)^* \\ &\quad - M_p^t((M_p^t)^* M_p^t)^{-1} (M_p^t)^* Y_p^{t_1}((Q_p^{t_1})^* Q_p^{t_1})^{-1} (Q_p^{t_1})^*. \end{aligned}$$

*Proof.* To simplify the notation, we drop the superscript  $t$  or  $t_1$  in the following proof. It should be understood that  $Q_p$  represents  $Q_p^{t_1}$ ,  $Y_p$  represents  $Y_p^{t_1}$ ,  $M_p$  represents  $M_p^t$ , and  $X_p$  represents  $X_p^t$ .

First let us consider projections of a deterministic matrix. Let  $\hat{A}_p$  be a deterministic matrix that satisfies the linear constraints  $Y_p = \hat{A}_p Q_p$  and  $X_p = \hat{A}_p^* M_p$ , then we have

$$\begin{aligned} \hat{A}_p &= \hat{A}_p Q_p (Q_p^* Q_p)^{-1} Q_p^* + \hat{A}_p (\mathbf{I} - Q_p (Q_p^* Q_p)^{-1} Q_p^*), \\ \hat{A}_p &= M_p (M_p^* M_p)^{-1} M_p^* \hat{A}_p + (\mathbf{I} - M_p (M_p^* M_p)^{-1} M_p^*) \hat{A}_p. \end{aligned}$$

Combining the two equations above, as well as the two linear constraints, we can write

$$\hat{A}_p = Y_p (Q_p^* Q_p)^{-1} Q_p^* + M_p (M_p^* M_p)^{-1} X_p - M_p (M_p^* M_p)^{-1} M_p^* Y_p (Q_p^* Q_p)^{-1} Q_p^* + \mathbf{P}_{M_p}^\perp \hat{A}_p \mathbf{P}_{Q_p}^\perp. \quad (26)$$

We now demonstrate the conditional distribution of  $A_1, \dots, A_P$ . Let  $S_1, \dots, S_P$  be arbitrary Borel sets on  $\mathbb{R}^{n \times N_1}, \dots, \mathbb{R}^{n \times N_P}$ , respectively.

$$\begin{aligned} &P(A_1 \in S_1, \dots, A_P \in S_P | A_p Q_p = Y_p, A_p^* M_p = X_p, \forall p) \\ &\stackrel{(a)}{=} P\left(\mathbf{E}_1^{t_1, t} + \mathbf{P}_{M_1^t}^\perp A_1 \mathbf{P}_{Q_1^{t_1}}^\perp \in S_1, \dots, \mathbf{E}_P^{t_1, t} + \mathbf{P}_{M_P^t}^\perp A_P \mathbf{P}_{Q_P^{t_1}}^\perp \in S_P | A_p Q_p = Y_p, A_p^* M_p = X_p, \forall p\right) \\ &\stackrel{(b)}{=} P\left(\mathbf{E}_1^{t_1, t} + \mathbf{P}_{M_1^t}^\perp A_1 \mathbf{P}_{Q_1^{t_1}}^\perp \in S_1, \dots, \mathbf{E}_P^{t_1, t} + \mathbf{P}_{M_P^t}^\perp A_P \mathbf{P}_{Q_P^{t_1}}^\perp \in S_P\right) \\ &= P\left(\mathbf{E}_1^{t_1, t} + \mathbf{P}_{M_1^t}^\perp A_1 \mathbf{P}_{Q_1^{t_1}}^\perp \in S_1\right) \dots P\left(\mathbf{E}_P^{t_1, t} + \mathbf{P}_{M_P^t}^\perp A_P \mathbf{P}_{Q_P^{t_1}}^\perp \in S_P\right), \end{aligned} \quad (27)$$

which implies the desired result. In step (a),

$$\mathbf{E}_p^{t_1, t} = Y_p (Q_p^* Q_p)^{-1} Q_p^* + M_p (M_p^* M_p)^{-1} X_p^* - M_p (M_p^* M_p)^{-1} M_p^* Y_p (Q_p^* Q_p)^{-1} Q_p^*, \quad p = 1, \dots, P,$$



which follows from (26). Step (b) holds since  $\mathbf{P}_{M_p}^\perp A_p \mathbf{P}_{Q_p}^\perp$  is independent of the conditioning. The independence is demonstrated as follows. Notice that  $A_p Q_p = A_p \mathbf{P}_{Q_p}^\parallel Q_p$ . In what follows, we will show that  $A_p^\parallel := A_p \mathbf{P}_{Q_p}^\parallel$  is independent of  $A_r^\perp := A_r \mathbf{P}_{Q_r}^\perp$ , for  $p, r = 1, \dots, P$ . Then similar approach can be used to demonstrate that  $\mathbf{P}_{M_p}^\perp A_p$  is independent of  $\mathbf{P}_{M_r}^\parallel A_r$ . Together they provide the justification for step (b). Note that  $A_p^\parallel$  and  $A_r^\perp$  are jointly normal, hence it is enough to show they are uncorrelated.

$$\begin{aligned} \mathbb{E} \left\{ [A_p^\parallel]_{i,j} [A_r^\perp]_{m,l} \right\} &= \mathbb{E} \left\{ \left( \sum_{k=1}^N [A_p]_{i,k} [\mathbf{P}_{Q_p}^\parallel]_{k,j} \right) \left( \sum_{k=1}^N [A_r]_{m,k} (\mathbf{I}_{k,l} - [\mathbf{P}_{Q_r}^\parallel]_{k,l}) \right) \right\} \\ &\stackrel{(a)}{=} \frac{1}{n} \delta_0(i, m) \delta_0(p, r) \left( \sum_{k=1}^N [\mathbf{P}_{Q_p}^\parallel]_{k,j} \mathbf{I}_{k,l} - \sum_{k=1}^N [\mathbf{P}_{Q_p}^\parallel]_{k,j} [\mathbf{P}_{Q_r^\perp}^\parallel]_{k,l} \right) \\ &\stackrel{(b)}{=} \frac{1}{n} \delta_0(i, m) \delta_0(p, r) \left( [\mathbf{P}_{Q_p}^\parallel]_{l,j} - \sum_{k=1}^N [\mathbf{P}_{Q_p}^\parallel]_{k,j} [\mathbf{P}_{Q_r}^\parallel]_{l,k} \right) \stackrel{(c)}{=} 0, \end{aligned}$$

where  $\delta_0(i, j)$  is the Kronecker delta function. In the above, step (a) holds since the original matrix  $A$  has  $\mathcal{N}(0, 1/n)$  i.i.d. entries, step (b) holds since projectors are symmetric matrices, and step (c) follows the property of projectors  $\mathbf{P}^2 = \mathbf{P}$ .  $\square$

Combining the results in Lemma 2 and [9, Lemma 4], we have the following conditional distribution lemma.

**Lemma 3.** *For the vectors  $h_p^{t+1}$  and  $b_p^t$  defined in (11), the following holds for  $t \geq 1$ ,  $p = 1, \dots, P$ :*

$$b_p^0 | \mathcal{S}^{0,0} \stackrel{d}{=} (\sigma_p^0)^\perp Z_p'^0 + \Delta_p^{0,0}, \quad h_p^1 | \mathcal{S}^{1,0} \stackrel{d}{=} (\tau_p^0)^\perp Z_p^0 + \Delta_p^{1,0}, \quad (28)$$

$$b_p^t | \mathcal{S}^{t,t} \stackrel{d}{=} \sum_{i=0}^{t-1} \hat{\gamma}_{p,i}^t b_p^i + (\sigma_p^t)^\perp Z_p'^t + \Delta_p^{t,t}, \quad h_p^t | \mathcal{S}^{t+1,t} \stackrel{d}{=} \sum_{i=0}^{t-1} \hat{\alpha}_{p,i}^t h_p^{i+1} + (\tau_p^t)^\perp Z_p^t + \Delta_p^{t+1,t}, \quad (29)$$

where

$$\Delta_p^{0,0} = \left( \frac{\|(q_p^0)^\perp\|}{\sqrt{n}} - (\sigma_p^0)^\perp \right) Z_p'^0 \quad (30)$$

$$\Delta_p^{1,0} = \left[ \left( \frac{\|(m_p^0)^\perp\|}{\sqrt{n}} - (\tau_p^0)^\perp \right) \mathbf{I} - \frac{\|(m_p^0)^\perp\|}{\sqrt{n}} \mathbf{P}_{q_p^0}^\parallel \right] Z_p^0 + q_p^0 \left( \frac{\|q_p^0\|^2}{n} \right)^{-1} \left( \frac{(b_p^0)^* (m_p^0)^\perp}{n} - \frac{\|q_p^0\|^2}{n} \right) \quad (31)$$

$$\begin{aligned} \Delta_p^{t,t} &= \sum_{i=0}^{t-1} (\gamma_{p,i}^t - \hat{\gamma}_{p,i}^t) b_p^i + \left[ \left( \frac{\|(q_p^t)^\perp\|}{\sqrt{n}} - (\sigma_p^t)^\perp \right) \mathbf{I} - \frac{\|(q_p^t)^\perp\|}{\sqrt{n}} \mathbf{P}_{M_p^t}^\parallel \right] Z_p'^t \\ &\quad + M_p^t \left( \frac{(M_p^t)^* M_p^t}{n} \right)^{-1} \left( \frac{(H_p^t)^* (q_p^t)^\perp}{n} - \frac{(M_p^t)^*}{n} \left[ \lambda_p^t m_p^{t-1} - \sum_{i=1}^{t-1} \lambda_{p,i}^t \gamma_{p,i}^t m_p^{i-1} \right] \right) \end{aligned} \quad (32)$$

$$\begin{aligned} \Delta_p^{t+1,t} &= \sum_{i=0}^{t-1} (\alpha_{p,i}^t - \hat{\alpha}_{p,i}^t) h_p^{i+1} + \left[ \left( \frac{\|(m_p^t)^\perp\|}{\sqrt{n}} - (\tau_p^t)^\perp \right) \mathbf{I} - \frac{\|(m_p^t)^\perp\|}{\sqrt{n}} \mathbf{P}_{Q_p^{t+1}}^\parallel \right] Z_p^t \\ &\quad + Q_p^{t+1} \left( \frac{(Q_p^{t+1})^* Q_p^{t+1}}{n} \right)^{-1} \left( \frac{(B_p^{t+1})^* (m_p^t)^\perp}{n} - \frac{(Q_p^{t+1})^*}{n} \left[ q_p^t - \sum_{i=0}^{t-1} \alpha_{p,i}^t q_p^i \right] \right), \end{aligned} \quad (33)$$

where  $Z_p'^t \in \mathbb{R}^n$  and  $Z_p^t \in \mathbb{R}^{N_p}$  are random vectors with independent standard normal elements, and are independent of the corresponding sigma algebras. Moreover,  $Z_p'^t$  is independent of  $Z_q'^t$  and  $Z_p^t$  is independent of  $Z_q^t$  when  $p \neq q$ .

*Proof.* The proof for each individual  $p \in [P]$  is similar to the proof for [9, Lemma 4]. The claim that  $Z_p'^t$  is independent of  $Z_q'^t$  and  $Z_p^t$  is independent of  $Z_q^t$  when  $p \neq q$  follows from Lemma 2, where we have that  $\tilde{A}_p$  is independent of  $\tilde{A}_q$  for  $p \neq q$ .  $\square$

### 3.4 Main Concentration Lemma

We use the shorthand  $X_n \doteq c$  to denote the concentration inequality  $P(|X_n - c| \geq \epsilon) \leq K_t e^{-\kappa_t n \epsilon}$ . As specified in the theorem statement, the lemma holds for all  $\epsilon \in (0, 1)$ , with  $K_t, \kappa_t$  denoting the generic constants dependent on  $t$ , but not on  $n, \epsilon$ .

**Lemma 4.** *With the  $\doteq$  notation defined above, the following holds for all  $t \geq 0$ ,  $p = 1, \dots, P$ .*

(a)

$$P\left(\frac{\|\Delta_p^{t,t}\|^2}{n} \geq \epsilon\right) \leq K_t e^{-\kappa_t n \epsilon}. \quad (34)$$

$$P\left(\frac{\|\Delta_p^{t+1,t}\|^2}{n} \geq \epsilon\right) \leq K_t e^{-\kappa_t n \epsilon}. \quad (35)$$

(b) (i) For pseudo Lipschitz functions  $\phi_h : \mathbb{R}^{t+2} \rightarrow \mathbb{R}$ ,

$$\frac{1}{N_p} \sum_{i=1}^{N_p} \phi_h(h_{p,i}^1, \dots, h_{p,i}^{t+1}, x_{p,i}) \doteq \mathbb{E}[\phi_h(\tau_p^0 \tilde{Z}_p^0, \dots, \tau_p^t \tilde{Z}_p^t, X)]. \quad (36)$$

(ii) Let  $\psi_h : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a bounded function that is differentiable in the first argument except possibly at a finite number of points, with bounded derivative when it exists. Then,

$$\frac{1}{N_p} \sum_{i=1}^{N_p} \psi_h(h_{p,i}^{t+1}, x_{p,i}) \doteq \mathbb{E}[\psi_h(\tau_p^t \tilde{Z}_p^t, X)], \quad (37)$$

where  $\{\tilde{Z}_p^t\}$  is defined in (20), and  $X \sim p_X$  is independent of  $\{\tilde{Z}_p^t\}$ .

(iii) For pseudo-Lipschitz function  $\phi_{h_b} : \mathbb{R}^{P(t+1)+1} \rightarrow \mathbb{R}$ ,

$$\frac{1}{n} \sum_{i=1}^n \phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^t, \dots, b_{P,i}^t, w_i) \doteq \mathbb{E}\left[\phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, \dots, \sigma_1^t \check{Z}_1^t, \dots, \sigma_P^t \check{Z}_P^t, W)\right], \quad (38)$$

where  $\{\check{Z}_p^t\}$  is defined in (20), and  $W \sim p_W$  is independent of  $\{\check{Z}_p^t\}$ .

(c)

$$\frac{(h_p^{t+1})^* q_p^0}{n} \doteq 0, \quad \frac{(h_p^{t+1})^* x_p}{n} \doteq 0. \quad (39)$$

$$\frac{(b_p^t)^* w}{n} \doteq 0. \quad (40)$$

(d) For all  $0 \leq r \leq t$ ,  $q \neq p$ ,

$$\frac{(h_p^{r+1})^* h_p^{t+1}}{n} \doteq \check{E}_p^{r,t}. \quad (41)$$

$$\frac{(b_p^r)^* b_p^t}{n} \doteq \tilde{E}_p^{r,t}. \quad (42)$$

(e) For all  $0 \leq r \leq t$ ,

$$\frac{(q_p^0)^* q_p^{t+1}}{n} \doteq \tilde{E}_p^{r+1,t+1}, \quad \frac{(q_p^{r+1})^* q_p^{t+1}}{n} \doteq \tilde{E}_p^{r+1,t+1}. \quad (43)$$

$$\frac{(m_p^r)^* m_p^t}{n} \doteq \check{E}_p^{r,t}. \quad (44)$$

(f) Define  $\hat{\lambda}_p^{t+1} = \delta_p^{-1} \mathbb{E}[f'_t(\tau_p^t \tilde{Z}_p^t, X)]$ . For all  $0 \leq r \leq t$ ,

$$\lambda_p^t \doteq \hat{\lambda}_p^t, \quad \frac{(h_p^{r+1})^* q_p^{t+1}}{n} \doteq \hat{\lambda}_p^{t+1} \check{E}_p^{r,t}, \quad \frac{(h_p^{t+1})^* q_p^{r+1}}{n} \doteq \hat{\lambda}_p^{r+1} \check{E}_p^{r,t}. \quad (45)$$

$$\frac{(b_p^r)^* m_p^t}{n} \doteq \tilde{E}_p^{r,t}, \quad \frac{(b_p^t)^* m_p^r}{n} \doteq \tilde{E}_p^{r,t}. \quad (46)$$

$$(47)$$

(g) For  $\mathbf{Q}_p^{t+1} = \frac{1}{n} (Q_p^{t+1})^* Q_p^{t+1}$  and  $\mathbf{M}_p^t = \frac{1}{n} (M_p^t)^* M_p^t$ , when the inverses exist, for all  $0 \leq i, j \leq t$ ,  $0 \leq i', j' \leq t-1$ ,

$$[(\mathbf{Q}_p^{t+1})^{-1}]_{i+1,j+1} \doteq [(\tilde{C}_p^{t+1})^{-1}]_{i+1,j+1}, \quad \gamma_{p,i}^{t+1} \doteq \hat{\gamma}_{p,i}^{t+1}. \quad (48)$$

$$[(\mathbf{M}_p^t)^{-1}]_{i'+1,j'+1} \doteq [(\check{C}_p^t)^{-1}]_{i'+1,j'+1}, \quad \alpha_{p,i'}^t \doteq \hat{\alpha}_{p,i'}^t, \quad (49)$$

where  $\hat{\gamma}_{p,i}^{t+1}$  and  $\hat{\alpha}_{p,i}^t$  are defined in (22).

(h)

$$\frac{\|(q_p^{t+1})_\perp\|^2}{n} \doteq (\sigma_p^{t+1})_\perp^2. \quad (50)$$

$$\frac{\|(m_p^t)_\perp\|^2}{n} \doteq (\tau_p^t)_\perp^2, \quad (51)$$

where  $(\sigma_p^{t+1})_\perp^2$  and  $(\tau_p^t)_\perp^2$  are defined in (23).

### 3.5 Proof of Theorem 1

Applying Part (b)(i) of Lemma 4, the proof the Theorem 1 is the same as the proof in [9] and therefore is not repeated here.

## 4 Proof of Lemma 4

The proof of Lemma 4 uses induction in the iteration index  $t$ . We label as  $\mathcal{B}_t$  the results (35), (38), (40), (42), (44), (46), (49), (51). The rest of the results in Lemma 4 are labels as  $\mathcal{H}_{t+1}$ . In the proof,  $K, \kappa$  are used as universal constant in the upper bounds.

#### 4.1 Step 1: Showing $\mathcal{B}_0$ holds.

In the following, when a statement involves  $p = 1, \dots, P$ , we demonstrate the proof for  $p = 1$ , and the cases when  $p = 2, \dots, P$  can be obtained similarly; unless otherwise specified.

(a) Recall the definition of  $\Delta_1^{0,0}$  in (30), we have

$$\begin{aligned} P\left(\frac{\|\Delta_1^{0,0}\|}{n} \geq \epsilon\right) &\leq P\left(\left|\frac{\|q_1^0\|}{\sqrt{n}} - (\sigma_1^0)_\perp\right| \geq \sqrt{\frac{\epsilon}{2}}\right) + P\left(\left|\frac{\|Z_1'^0\|}{\sqrt{n}} - 1\right| \geq \sqrt{\frac{\epsilon}{2}}\right) \\ &\stackrel{(a)}{\leq} Ke^{-\kappa n \epsilon} + Ke^{-\kappa n \epsilon}, \end{aligned} \quad (52)$$

where step (a) uses our assumption on  $q_1^0$  in (13), Lemma B.2, and Lemma A.2.

(b) (iii) Recall the conditional distribution of  $b_p^0$  in (28), by Lemma B.1,

$$\begin{aligned} &P\left(\left|\frac{1}{n} \sum_{i=1}^n \phi_b(b_{1,i}^0, \dots, b_{P,i}^0, w_i) - \mathbb{E}\left[\phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, W)\right]\right| \geq \epsilon\right) \\ &\leq P\left(\frac{1}{n} \sum_{i=1}^n \left|\phi_b(\sigma_1^0 Z_{1,i}'^0 + \Delta_{1,i}^{0,0}, \dots, \sigma_P^0 Z_{P,i}'^0 + \Delta_{P,i}^{0,0}, w_i) - \phi_b(\sigma_1^0 Z_{1,i}'^0, \dots, \sigma_P^0 Z_{P,i}'^0, w_i)\right| \geq \frac{\epsilon}{2}\right) \\ &+ P\left(\left|\frac{1}{n} \sum_{i=1}^n \phi_b(\sigma_1^0 Z_{1,i}'^0, \dots, \sigma_P^0 Z_{P,i}'^0, w_i) - \mathbb{E}\left[\phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, W)\right]\right| \geq \frac{\epsilon}{2}\right). \end{aligned}$$

Label the two terms above as  $T_1$  and  $T_2$ , we will show that both are upper bounded by  $Ke^{-\kappa n \epsilon^2}$ . First consider  $T_1$ .

$$\begin{aligned} T_1 &\stackrel{(a)}{\leq} P\left(\frac{1}{n} \sum_{i=1}^n \left(1 + \sqrt{\sum_{u=1}^P (\sigma_u^0 Z_{u,i}'^0 + \Delta_{u,i}^{0,0})^2} + \sqrt{\sum_{u=1}^P (\sigma_u^0 Z_{u,i}'^0)^2}\right) \left(\sqrt{\sum_{u=1}^P (\Delta_{u,i}^{0,0})^2}\right) \geq \frac{\epsilon}{2L}\right) \\ &\stackrel{(b)}{\leq} P\left(\sqrt{1 + \sum_{u=1}^P \frac{\|\Delta_u^{0,0}\|^2}{n}} + 4 \sum_{u=1}^P (\sigma_u^0)^2 \frac{\|Z_u'^0\|^2}{n} \sqrt{\sum_{u=1}^P \frac{\|\Delta_u^{0,0}\|^2}{n}} \geq \frac{\epsilon}{2\sqrt{3}L}\right) \\ &\leq P\left(\left(1 + \sum_{u=1}^P \frac{\|\Delta_u^{0,0}\|}{\sqrt{n}} + 2 \sum_{u=1}^P \sigma_u^0 \frac{\|Z_u'^0\|}{\sqrt{n}}\right) \left(\sum_{u=1}^P \frac{\|\Delta_u^{0,0}\|}{\sqrt{n}}\right) \geq \frac{\epsilon}{2\sqrt{3}L}\right). \end{aligned} \quad (53)$$

In the above, step (a) holds because  $\phi_b$  is pseudo-Lipschitz. Step (b) uses triangle inequality to obtain  $\sqrt{\sum_{u=1}^P (\sigma_u^0 Z_{u,i}'^0 + \Delta_{u,i}^{0,0})^2} \leq \sqrt{\sum_{u=1}^P (\sigma_u^0 Z_{u,i}'^0)^2} + \sqrt{\sum_{u=1}^P (\Delta_{u,i}^{0,0})^2}$  (think of it as  $\ell_2$ -norm of length- $P$  vectors), followed by Cauchy-Schwarz and Lemma B.5.

From (53), we have

$$\begin{aligned} T_1 &\leq \sum_{u=1}^P P\left(\frac{\|Z_u'^0\|}{\sqrt{n}} - 1 \geq 1\right) + P\left(\sum_{u=1}^P \frac{\|\Delta_u^{0,0}\|}{\sqrt{n}} \geq \frac{\epsilon \min\{1, \frac{1}{2\sqrt{3}L}\}}{2 + 8 \sum_{u=1}^P \sigma_u^0}\right) \\ &\stackrel{(a)}{\leq} \sum_{u=1}^P P\left(\frac{\|Z_u'^0\|}{\sqrt{n}} - 1 \geq 1\right) + \sum_{u=1}^P P\left(\frac{\|\Delta_u^{0,0}\|}{\sqrt{n}} \geq \frac{\epsilon \min\{1, \frac{1}{2\sqrt{3}L}\}}{P(2 + 8 \sum_{u=1}^P \sigma_u^0)}\right) \\ &\stackrel{(b)}{\leq} Ke^{-n} + Ke^{-\kappa n \epsilon}, \end{aligned}$$

where step (a) follows from Lemma B.1 and step (b) uses Lemma A.2 and  $\mathcal{B}_0(a)$ .

Next consider  $T_2$ . Using Lemma B.1, we have

$$T_2 \leq P \left( \left| \frac{1}{n} \sum_{i=1}^n \left( \phi_b(\sigma_1^0 Z'_{1,i}, \dots, \sigma_P^0 Z'_{P,i}, w_i) - \mathbb{E}_{(Z_{1,i}, \dots, Z_{P,i})} \left[ \phi_b(\sigma_1^0 Z'_{1,i}, \dots, \sigma_P^0 Z'_{P,i}, w_i) \right] \right) \right| \geq \frac{\epsilon}{4} \right) \\ + P \left( \left| \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{(Z_{1,i}, \dots, Z_{P,i})} \left[ \phi_b(\sigma_1^0 Z'_{1,i}, \dots, \sigma_P^0 Z'_{P,i}, w_i) \right] - \mathbb{E} \left[ \phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, W) \right] \right| \geq \frac{\epsilon}{4} \right).$$

Label the two terms above as  $T_{2,a}$  and  $T_{2,b}$ .  $T_{2,a}$  has the desired bound by first noticing that the function  $\tilde{\phi}_{b,i} : \mathbb{R}^P \rightarrow \mathbb{R}$  defined as  $\tilde{\phi}_{b,i}(x_1, \dots, x_P) := \phi_b(x_1, \dots, x_P, w_i)$  is PL(2) since  $\phi_b$  is, and then applying Lemma A.4 to sequence of i.i.d. length- $P$  random vectors consisting of i.i.d.  $\mathcal{N}(0, 1)$  Gaussian random variables (variance factor  $\nu = 1$  in Lemma A.4).  $T_{2,b}$  has the desired bound by first noticing that the function  $\hat{\phi}_b : \mathbb{R} \rightarrow \mathbb{R}$  defined as  $\hat{\phi}_b(x) := \mathbb{E} \left[ \phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, x) \right]$  is PL(2) by Lemma C.2.

(c) The function  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, w_i) := b_{1,i}^0 w_i$  is PL(2) by Lemma C.1. Apply  $\mathcal{B}_0(b)(iii)$ , we have

$$\frac{1}{n} \sum_{i=1}^n b_{1,i}^0 w_i \doteq \mathbb{E} \left[ \sigma_1^0 \check{Z}_1^0 W \right] = 0.$$

(d) The function  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, w_i) := (b_{1,i}^0)^2$  is PL(2) by Lemma C.1. Apply  $\mathcal{B}_0(b)(iii)$ , we have

$$\frac{1}{n} \sum_{i=1}^n (b_{1,i}^0)^2 \doteq \mathbb{E} \left[ (\sigma_1^0 \check{Z}_1^0)^2 \right] = \tilde{E}_1^{0,0},$$

where the last equality uses (20).

(e) Recall that  $m_1^0 = b_1^0 - w$ . The function  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, w_i) := b_{1,i}^0 - w_i$  is PL(2). Apply  $\mathcal{B}_0(b)(iii)$ , we have

$$\frac{(m_1^0)^* m_1^0}{n} = \frac{1}{n} \sum_{i=1}^n (b_{1,i}^0 - w_i)^2 \doteq \mathbb{E} \left[ (\sigma_1^0 \check{Z}_1^0 - W)^2 \right] = (\sigma_1^0)^2 + \sigma_W^2 = \tilde{E}_1^{0,0}.$$

The last equality follows from (21) by noticing that  $\tilde{E}_1^{\theta(0), \theta(0)} = 0$  by definition.

(f) The function  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, w_i) := b_{1,i}^0 (b_{1,i}^0 - w_i)$  is PL(2) by Lemma C.1. Apply  $\mathcal{B}_0(b)(iii)$ , we have

$$\frac{(b_1^0)^* m_1^0}{n} = \frac{1}{n} b_{1,i}^0 (b_{1,i}^0 - w_i) \doteq \mathbb{E} \left[ \sigma_1^0 \check{Z}_1^0 (\sigma_1^0 \check{Z}_1^0 - W) \right] = \tilde{E}_1^{0,0}.$$

(g) Nothing to prove here.

(h) The result is equivalent to  $\mathcal{B}_0(e)$ , since  $\|(m_p^0)_\perp\| = \|m_1^0\|$  and  $(\tau_p^0)_\perp^2 = (\tau_p^0)^2$ .

## 4.2 Step 2: Showing $\mathcal{H}_1$ holds.

The proof is the same as in [9] and is not repeated here.

## 4.3 Step 3: Showing $\mathcal{B}_t$ holds.

We prove the statements in  $\mathcal{B}_t$  assuming that  $\mathcal{B}_{t-1}$  and  $\mathcal{H}_t$  holds by inductive hypothesis.

(a) The proof is the same as in [9].

(b)(iii) For brevity, define  $\mathbb{E}_{\phi_b} := \mathbb{E} \left[ \phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, \dots, \sigma_1^t \check{Z}_1^t, \dots, \sigma_P^t \check{Z}_P^t, W) \right]$ , and

$$\begin{aligned} a_i &= (b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t b_{1,i}^r + (\sigma_1^t)_\perp Z_{1,i}'^t + [\Delta_1^{t,t}]_i, \dots, \sum_{r=0}^{t-1} \hat{\gamma}_{P,r}^t b_{P,i}^r + (\sigma_P^t)_\perp Z_{P,i}'^t + [\Delta_P^{t,t}]_i, w_i) \\ c_i &= (b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t b_{1,i}^r + (\sigma_1^t)_\perp Z_{1,i}'^t, \dots, \sum_{r=0}^{t-1} \hat{\gamma}_{P,r}^t b_{P,i}^r + (\sigma_P^t)_\perp Z_{P,i}'^t, w_i), \end{aligned}$$

for  $i = 1, \dots, n$ . Hence,  $a, c$  are length- $n$  vectors with elements in  $\mathbb{R}^{P(t+1)+1}$ .

Then, using the conditional distribution of  $b_p^t$  defined in Lemma 3 (29), together with Lemma B.1, we have

$$\begin{aligned} &P \left( \left| \frac{1}{n} \sum_{i=1}^n \phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^t, \dots, b_{P,i}^t, w_i) - \mathbb{E}_{\phi_b} \right| \geq \epsilon \right) \\ &\leq P \left( \left| \frac{1}{n} \sum_{i=1}^n (\phi_b(a_i) - \phi_b(c_i)) \right| \geq \frac{\epsilon}{2} \right) + P \left( \left| \frac{1}{n} \sum_{i=1}^n \phi_b(c_i) - \mathbb{E}_{\phi_b} \right| \geq \frac{\epsilon}{2} \right). \end{aligned}$$

Label the two terms above as  $T_1$  and  $T_2$ . To complete the proof, we show that both  $T_1$  and  $T_2$  are bounded above by  $Ke^{-\kappa n \epsilon^2}$ . First consider  $T_1$ . For convenience, we define  $\underline{\Delta}_i^{t,t} := ([\Delta_1^{t,t}]_i, \dots, [\Delta_P^{t,t}]_i)$ , for  $i = 1, \dots, n$ .

$$\begin{aligned} T_1 &\stackrel{(a)}{\leq} P \left( \frac{1}{n} \sum_{i=1}^n (1 + \|a_i\| + \|c_i\|) \|a_i - c_i\| \geq \frac{\epsilon}{2L} \right) \stackrel{(b)}{\leq} P \left( \frac{1}{n} \sum_{i=1}^n (1 + \|\underline{\Delta}_i^{t,t}\| + 2\|c_i\|) \|\underline{\Delta}_i^{t,t}\| \geq \frac{\epsilon}{2L} \right) \\ &\stackrel{(c)}{\leq} P \left( \sqrt{1 + \sum_{u=1}^P \frac{\|\Delta_u^{t,t}\|^2}{n}} + 4 \frac{\|c\|^2}{n} \sqrt{\sum_{u=1}^P \frac{\|\Delta_u^{t,t}\|^2}{n}} \geq \frac{\epsilon}{2\sqrt{3}L} \right) \\ &\leq P \left( \left( 1 + \sum_{u=1}^P \frac{\|\Delta_u^{t,t}\|}{\sqrt{n}} + 2 \frac{\|c\|}{\sqrt{n}} \right) \left( \sum_{u=1}^P \frac{\|\Delta_u^{t,t}\|}{\sqrt{n}} \right) \geq \frac{\epsilon}{2\sqrt{3}L} \right), \end{aligned}$$

where step (a) holds because  $\phi_b$  is PL(2), step (b) holds because  $\|a_i\| \leq \|c_i\| + \|\underline{\Delta}_i^{t,t}\|$  and  $\|a_i - c_i\| = \|\underline{\Delta}_i^{t,t}\|$ , and step (c) holds by Cauchy-Schwarz followed by Lemma B.5. Further,  $\|c\|$  can be bounded as follows by applying Lemma B.5,

$$\|c\|^2 \leq \sum_{u=1}^P \sum_{r=0}^{t-1} \|b_u^r\|^2 + 2 \sum_{u=1}^P \sum_{r=0}^{t-1} \sum_{l=0}^{t-1} \hat{\gamma}_{u,r}^t \hat{\gamma}_{u,l}^t (b_u^r)^* b_u^l + 2 \sum_{u=0}^P (\sigma_u^t)_\perp^2 \|Z_u'^t\|^2 + \|w\|^2.$$

Denote the RHS of the above by  $\tilde{c}$ . We will show that  $\frac{1}{n} \tilde{c}$  concentrates to  $\mathbb{E}_{\tilde{c}}$  defined as

$$\mathbb{E}_{\tilde{c}} := \sum_{u=1}^P \sum_{r=0}^{t-1} \tilde{E}_u^{r,r} + 2 \sum_{u=1}^P \sum_{r=0}^{t-1} \sum_{l=0}^{t-1} \hat{\gamma}_{u,r}^t \hat{\gamma}_{u,l}^t \tilde{E}_u^{r,l} + 2 \sum_{u=1}^P (\sigma_u^t)_\perp^2 + \sigma_W^2 = \sum_{u=1}^P \left( (\sigma_u^t)^2 + 2(\sigma_u^t)^2 \right) + \sigma_W^2.$$

In the above, the last inequality uses the definitions in Section 3.2 and follows from

$$\begin{aligned} \sum_{r=0}^{t-1} \sum_{l=0}^{t-1} \hat{\gamma}_{u,r}^t \hat{\gamma}_{u,l}^t \tilde{E}_u^{r,l} &= (\hat{\gamma}_u^t)^* \tilde{C}_u^t \hat{\gamma}_u^t = [(\tilde{E}_u^t)^* (\tilde{C}_u^t)^{-1}] \tilde{C}_u^t [(\tilde{C}_u^t)^{-1} \tilde{E}_u^t] = (\tilde{E}_u^t)^* (\tilde{C}_u^t)^{-1} \tilde{E}_u^t = (\sigma_u^t)^2 - (\sigma_u^t)_\perp^2. \end{aligned} \tag{54}$$

To see the concentration, let  $\mathbf{a} = Pt + t^2 + P + 1$ ,

$$\begin{aligned} P \left( \left| \frac{1}{n} \tilde{\mathbf{c}} - \mathbb{E}_{\tilde{\mathbf{c}}} \right| \geq \epsilon' \right) &\stackrel{(a)}{\leq} \sum_{u=1}^P \sum_{r=0}^{t-1} P \left( \left| \frac{\|b_u^r\|^2}{n} - \tilde{E}_u^{r,r} \right| \geq \frac{\epsilon'}{\mathbf{a}} \right) + \sum_{u=1}^P \sum_{r=0}^{t-1} \sum_{l=0}^{t-1} P \left( \left| \frac{(b_u^r)^* b_u^l}{n} - \tilde{E}_u^{r,l} \right| \geq \frac{\epsilon'}{2\mathbf{a} \hat{\gamma}_{u,r}^t \hat{\gamma}_{u,l}^t} \right) \\ &+ P \left( \left| \frac{\|Z_u'^t\|^2}{n} - 1 \right| \geq \frac{\epsilon'}{2\mathbf{a}(\sigma_u^t)_{\perp}^2} \right) + P \left( \left| \frac{\|w\|^2}{n} - \sigma_W^2 \right| \geq \frac{\epsilon'}{\mathbf{a}} \right) \stackrel{(b)}{\leq} K e^{-\kappa n \epsilon'^2}, \end{aligned} \quad (55)$$

where step (a) follows from Lemma B.1 and step (b) follows from the inductive hypothesis  $\mathcal{B}_0(d) - \mathcal{B}_{t-1}(d)$ , Lemma A.2, and the assumption on  $w$ . Using this result,  $T_1$  can be bounded as follows,

$$\begin{aligned} T_1 &\leq P \left( \left( 1 + \sum_{u=1}^P \frac{\|\Delta_u^{t,t}\|}{\sqrt{n}} + 2 \left( \frac{\tilde{\mathbf{c}}}{\sqrt{n}} - \mathbb{E}_{\tilde{\mathbf{c}}} \right) + 2\mathbb{E}_{\tilde{\mathbf{c}}} \right) \left( \sum_{u=1}^P \frac{\|\Delta_u^{t,t}\|}{\sqrt{n}} \right) \geq \frac{\epsilon}{2\sqrt{3}L} \right) \\ &\stackrel{(a)}{\leq} P \left( \left| \frac{\tilde{\mathbf{c}}}{\sqrt{n}} - \mathbb{E}_{\tilde{\mathbf{c}}}^{1/2} \right| \geq \epsilon \right) + \sum_{u=1}^P P \left( \frac{\|\Delta_u^{t,t}\|}{\sqrt{n}} \geq \frac{\epsilon}{2\sqrt{3}LP(4 + 2\mathbb{E}_{\tilde{\mathbf{c}}}^{1/2})} \right) \stackrel{(b)}{\leq} K e^{-\kappa n \epsilon^2} + K e^{-\kappa n \epsilon^2}, \end{aligned}$$

where step (a) uses Lemma B.1 and step (b) uses (55), the results in  $\mathcal{B}_t(a)$ , and Lemma B.2.

Next consider  $T_2$ . Let  $Z = (Z_1, \dots, Z_P)$  be a vector of i.i.d.  $\mathcal{N}(0, 1)$  random variables, and it is independent of  $Z_u'^r$  for all  $r = 1, \dots, t-1$  and  $u = 1, \dots, P$ . For each  $i \in [n]$ , define the function  $\tilde{\phi}_{b,i} : \mathbb{R}^P \rightarrow \mathbb{R}$  as for  $z = (z_1, \dots, z_P) \in \mathbb{R}^P$ ,

$$\tilde{\phi}_{b,i}(z) := \phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t b_{1,i}^r + (\sigma_1^t)_{\perp} z_1, \dots, \sum_{r=0}^{t-1} \hat{\gamma}_{P,r}^t b_{P,i}^r + (\sigma_P^t)_{\perp} z_P, w_i).$$

Then  $\tilde{\phi}_{b,i}$  is PL(2) for all  $i \in [n]$ . For brevity, let  $Z'_i := (Z_{1,i}^t, \dots, Z_{P,i}^t)$ , for all  $i \in [n]$ , then

$$T_2 \leq P \left( \left| \frac{1}{n} \sum_{i=1}^n \left( \tilde{\phi}_{b,i}(Z'_i) - \mathbb{E}_Z[\tilde{\phi}_{b,i}(Z)] \right) \right| \geq \frac{\epsilon}{4} \right) + P \left( \left| \frac{1}{n} \sum_{i=1}^n \mathbb{E}_Z[\tilde{\phi}_{b,i}(Z)] - \mathbb{E}_{\phi_b} \right| \geq \frac{\epsilon}{4} \right).$$

Label the two terms in the above as  $T_1$  and  $T_2$ .  $T_1$  is bounded above by  $K e^{-\kappa n \epsilon^2}$  using Lemma A.4.  $T_2$  can be written as

$$T_2 = P \left( \left| \frac{1}{n} \sum_{i=1}^n \phi'_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, w_i) - \mathbb{E}_{\phi_b} \right| \geq \frac{\epsilon}{4} \right),$$

where the function  $\phi'_b : \mathbb{R}^{P(t+1)+1} \rightarrow \mathbb{R}$  is defined as

$$\begin{aligned} \phi'_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, w_i) &:= \mathbb{E}_Z[\tilde{\phi}_{b,i}(Z)] \\ &= \mathbb{E}_Z \left[ \phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t b_{1,i}^r + (\sigma_1^t)_{\perp} Z_1, \dots, \sum_{r=0}^{t-1} \hat{\gamma}_{P,r}^t b_{P,i}^r + (\sigma_P^t)_{\perp} Z_P, w_i) \right] \end{aligned}$$

Then  $\phi'_b$  is PL(2) by Lemma C.2. By inductive hypothesis  $\mathcal{B}_0(b)(iii) - \mathcal{B}_{t-1}(b)(iii)$ , we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \phi'_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^{t-1}, \dots, b_{P,i}^{t-1}, w_i) &\doteq \mathbb{E} \left[ \phi'_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, \dots, \sigma_1^{t-1} \check{Z}_1^{t-1}, \dots, \sigma_P^{t-1} \check{Z}_P^{t-1}, W) \right] \\ &= \mathbb{E} \left[ \phi_b(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, \dots, \sigma_1^{t-1} \check{Z}_1^{t-1}, \dots, \sigma_P^{t-1} \check{Z}_P^{t-1}, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t \sigma_1^r \check{Z}_1^r + (\sigma_1^t)_{\perp} Z_1, \dots, \sum_{r=0}^{t-1} \hat{\gamma}_{P,r}^t \sigma_P^r \check{Z}_P^r + (\sigma_P^t)_{\perp} Z_P, W) \right]. \end{aligned}$$

To obtain the desired result, we need to show

$$(\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, \dots, \sigma_1^{t-1} \check{Z}_1^{t-1}, \dots, \sigma_P^{t-1} \check{Z}_P^{t-1}, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t \sigma_1^r \check{Z}_1^r + (\sigma_1^t)_\perp Z_1, \dots, \sum_{r=0}^{t-1} \hat{\gamma}_{1,r}^t \sigma_P^r \check{Z}_P^r + (\sigma_P^t)_\perp Z_P) \quad (56)$$

$$\stackrel{d}{=} (\sigma_1^0 \check{Z}_1^0, \dots, \sigma_P^0 \check{Z}_P^0, \dots, \sigma_1^t \check{Z}_1^t, \dots, \sigma_P^t \check{Z}_P^t). \quad (57)$$

Notice that these are zero-mean jointly Gaussian random vectors, hence we only need to demonstrate that their covariance matrices are equal. For the coordinates on the diagonal, we need to show

$$\mathbb{E} \left[ \left( \sum_{r=0}^{t-1} \hat{\gamma}_{p,r}^r \sigma_p^r \check{Z}_p^r + (\sigma_p^t)_\perp Z_p \right)^2 \right] = \mathbb{E} \left[ (\sigma_p^t \check{Z}_p^t)^2 \right] = (\sigma_p^t)^2, \quad \forall p = 1, \dots, P. \quad (58)$$

This is true since

$$\mathbb{E} \left[ \left( \sum_{r=0}^{t-1} \hat{\gamma}_{p,r}^r \sigma_p^r \check{Z}_p^r + (\sigma_p^t)_\perp Z_p \right)^2 \right] \stackrel{(a)}{=} \sum_{r=0}^{t-1} \sum_{l=0}^{t-1} \hat{\gamma}_{p,r}^t \hat{\gamma}_{p,l}^t \tilde{E}_p^{r,l} + (\sigma_p^t)_\perp^2 \stackrel{(b)}{=} (\sigma_p^t)^2,$$

where step (a) follows from the definition in Section 3.2 and step (b) follows from (54). Next consider the off-diagonal terms. Notice that  $\check{Z}_p^r$  is independent of  $\check{Z}_p^l$  for all  $0 \leq r, l \leq t$  whenever  $p \neq q$  and also the i.i.d. random vector  $(Z_1, \dots, Z_P)$  is independent of  $\check{Z}_p^r$  for all  $1 \leq p \leq P, 0 \leq r \leq t$ . Hence, we only need to show for all  $0 \leq l \leq t-1$ ,

$$\mathbb{E} \left[ (\sigma_p^l \check{Z}_p^l) \left( \sum_{r=0}^{t-1} \hat{\gamma}_{p,r}^t \sigma_p^r \check{Z}_p^r + (\sigma_p^t)_\perp Z_p \right) \right] = \mathbb{E} \left[ (\sigma_p^l \check{Z}_p^l) (\sigma_p^t \check{Z}_p^t) \right] = \tilde{E}_p^{l,t}.$$

This is true since

$$\mathbb{E} \left[ (\sigma_p^l \check{Z}_p^l) \left( \sum_{r=0}^{t-1} \hat{\gamma}_{p,r}^t \sigma_p^r \check{Z}_p^r + (\sigma_p^t)_\perp Z_p \right) \right] \stackrel{(a)}{=} \sum_{r=0}^{t-1} \hat{\gamma}_{p,r}^t \tilde{E}_p^{l,t} \stackrel{(b)}{=} [\tilde{C}_p^t \hat{\gamma}_p^t]_{l+1} \stackrel{(c)}{=} \tilde{E}_p^{l,t},$$

where step (a), (b), and (c) uses definition for  $\tilde{E}_p^{l,r}$ ,  $\tilde{C}_p^t$ , and  $\hat{\gamma}_p^t$ , respectively, in Section 3.2. Hence, we have proved (57).

(c) Apply the result in  $\mathcal{B}_t(b)(iii)$  to PL(2) function  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^t, \dots, b_{P,i}^t, w_i) := b_{1,i}^t w_i$ , then we have

$$\frac{1}{n} \sum_{i=1}^n b_{1,i}^t w_i \doteq \mathbb{E} \left[ \sigma_1^t \check{Z}_1^t W \right] = 0.$$

(d) Apply the result in  $\mathcal{B}_t(b)(iii)$  to PL(2) function  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^t, \dots, b_{P,i}^t, w_i) := b_{1,i}^r b_{1,i}^t$ , then we have

$$\frac{1}{n} \sum_{i=1}^n b_{1,i}^r b_{1,i}^t \doteq \sigma_1^r \sigma_1^t \mathbb{E} \left[ \check{Z}_1^r \check{Z}_1^t \right] = \tilde{E}_1^{r,t}.$$

(e) Define the function  $g : \mathbb{R}^{P+1} \rightarrow \mathbb{R}$  as  $g(x) = \sum_{i=1}^P x_i - x_{P+1}$  for  $x = (x_1, \dots, x_{P+1}) \in \mathbb{R}^{P+1}$ . Notice that  $g$  is Lipschitz. Recall the definition of  $m_1^t$  in (11), we have  $m_{1,i}^t = g(b_{2,i}^{\theta(t)}, \dots, b_{P,i}^{\theta(t)}, b_{1,i}^t, w_i)$ . Notice that  $\theta(t)$  equals to some integer that is less than or equal to  $t$ . Therefore, the function defined



as  $\phi_b(b_{1,i}^0, \dots, b_{P,i}^0, \dots, b_{1,i}^t, \dots, b_{P,i}^t, w_i) := g(b_{2,i}^{\theta(r)}, \dots, b_{P,i}^{\theta(r)}, b_{1,i}^r, w_i)g(b_{2,i}^{\theta(t)}, \dots, b_{P,i}^{\theta(t)}, b_{1,i}^t, w_i)$  is PL(2) by Lemma C.1. Applying the result in  $\mathcal{B}_t(b)(iii)$ , we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n m_{1,i}^r m_{1,i}^t &\doteq \mathbb{E} \left[ g(\sigma_2^{\theta(r)} \check{Z}_2^{\theta(r)}, \dots, \sigma_P^{\theta(r)} \check{Z}_P^{\theta(r)}, \sigma_1^r \check{Z}_1^r, W) g(\sigma_2^{\theta(t)} \check{Z}_2^{\theta(t)}, \dots, \sigma_P^{\theta(t)} \check{Z}_P^{\theta(t)}, \sigma_1^t \check{Z}_1^t, W) \right] \\ &= \tilde{E}_1^{r,t} + \sum_{u=2}^P \tilde{E}_u^{\theta(r), \theta(t)} + \sigma_W^2 = \tilde{E}_1^{r,t}. \end{aligned}$$

(f) Using the function  $g$  defined above and applying  $\mathcal{B}(b)(iii)$ , we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n b_{1,i}^r m_{1,i}^t &= \frac{1}{n} \sum_{i=1}^n b_{1,i}^r g(b_{2,i}^{\theta(t)}, \dots, b_{P,i}^{\theta(t)}, b_{1,i}^t, w_i) \doteq \mathbb{E} \left[ \sigma_1^r \check{Z}_1^r g(\sigma_2^{\theta(t)} \check{Z}_2^{\theta(t)}, \dots, \sigma_P^{\theta(t)} \check{Z}_P^{\theta(t)}, \sigma_1^t \check{Z}_1^t, W) \right] \\ &= \mathbb{E} \left[ \sigma_1^r \check{Z}_1^r \left( \sum_{u=2}^P \sigma_u^{\theta(0)} \check{Z}_u^{\theta(0)} + \sigma_1^t \check{Z}_u^t - W \right) \right] = \tilde{E}_1^{r,t}. \end{aligned}$$

The proof for  $\frac{1}{n} \sum_{i=1}^n b_{1,i}^t m_{1,i}^r \doteq \tilde{E}_1^{r,t}$  is similar.

The proof for (g) and (h) is the same as in [9].

#### 4.4 Step 4: Showing $\mathcal{H}_{t+1}$ holds.

The proof is the same as in [9] and is not repeated hear.

## A Sub-Gaussian Concentration Lemmas

**Lemma A.1.** [9] For a standard Gaussian random variable  $Z$  and  $\epsilon > 0$ ,  $P(|Z| \geq \epsilon) \leq 2e^{-\frac{1}{2}\epsilon^2}$ .

**Lemma A.2.** [9,  $\chi^2$ -concentration] For  $Z_i$ ,  $i \in [n]$  that are i.i.d.  $\sim \mathcal{N}(0, 1)$ , and  $0 \leq \epsilon \leq 1$ ,

$$P \left( \left| \frac{1}{n} \sum_{i=1}^n Z_i^2 - 1 \right| \geq \epsilon \right) \leq 2e^{-n\epsilon^2/8}.$$

**Lemma A.3.** [16] Let  $X$  be a centered sub-Gaussian random variable with variance factor  $\nu$ , i.e.,  $\ln \mathbb{E}[e^{tX}] \leq \frac{t^2 \nu}{2}$ ,  $\forall t \in \mathbb{R}$ . Then  $X$  satisfies:

1. For all  $x > 0$ ,  $P(X > x) \vee P(X < -x) \leq e^{-\frac{x^2}{2\nu}}$ , for all  $x > 0$ .

2. For every integer  $k \geq 1$ ,

$$\mathbb{E}[X^{2k}] \leq 2(k!)(2\nu)^k \leq (k!)(4\nu)^k. \quad (59)$$

**Lemma A.4.** Let  $\{Z_i\}_{i \in [n]} \in \mathbb{R}^d$  be a sequence of i.i.d. random vectors on  $\mathbb{R}^d$ , where  $d$  is an integer and each  $Z_i$  has i.i.d. sub-Gaussian entries with variance factor  $\nu$ . The functions  $\{f_i\}_{i \in [n]} : \mathbb{R}^d \rightarrow \mathbb{R}$  are pseudo-Lipschitz of order 2 each with pseudo-Lipschitz constant  $L_i$ . Let  $L := \max_{i \in [n]} L_i$ . Then for all  $\epsilon \in (0, 1)$ , there exist  $K, \kappa > 0$  depending on  $d, L$  but not  $n, \epsilon$  such that

$$P \left( \left| \frac{1}{n} \sum_{i=1}^n (f_i(Z_i) - \mathbb{E}[f_i(Z_i)]) \right| \geq \epsilon \right) \leq K e^{-\kappa n \epsilon^2}. \quad (60)$$

*Proof.* In what follows, we prove the upper-tail bound

$$P\left(\frac{1}{n}\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)]) \geq \epsilon\right) \leq Ke^{-\kappa n\epsilon^2}. \quad (61)$$

The lower-tail bound can then be obtained similarly. Together they prove the desired result.

Using the Cramér-Chernoff method:

$$P\left(\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)]) \geq n\epsilon\right) \leq e^{-nr\epsilon}\mathbb{E}\left[\exp\left(r\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)])\right)\right], \quad \forall r > 0. \quad (62)$$

Next we will show that there exist  $\kappa' > 0$  such that the expectation in (62) is bounded by

$$\mathbb{E}\left[\exp\left(r\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)])\right)\right] \leq e^{\kappa'nr^2}, \quad \text{for } 0 < r \leq (5\sqrt{2}L(d\nu + 12d^2\nu^2)^{1/2})^{-1}. \quad (63)$$

Then plugging (63) into (62), and choosing the optimal  $r = \frac{\epsilon}{2\kappa'}$ , we can ensure that  $r$  falls within the effective region defined in (62) for all  $\epsilon \in (0, 1)$  by choosing  $\kappa'$  big enough.

Now we show (63). Notice that

$$\mathbb{E}\left[\exp\left(-r\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)])\right)\right] \stackrel{(a)}{\geq} \exp\left(-r\mathbb{E}\left[\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)])\right]\right) = 1, \quad (64)$$

where step (a) follows from Jensen's inequality. Let  $\{\tilde{Z}_i\}_{i \in [n]}$  be an independent copy of  $\{Z_i\}_{i \in [n]}$ . Then we have

$$\begin{aligned} & \mathbb{E}\left[\exp\left\{r\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)])\right\}\right] \cdot 1 \\ & \stackrel{(a)}{\leq} \mathbb{E}\left[\exp\left\{r\sum_{i=1}^n(f_i(Z_i) - \mathbb{E}[f_i(Z_i)])\right\}\right] \mathbb{E}\left[\exp\left\{-r\sum_{i=1}^n(f_i(\tilde{Z}_i) - \mathbb{E}[f_i(\tilde{Z}_i)])\right\}\right] \\ & \stackrel{(b)}{=} \prod_{i=1}^n \mathbb{E}\left[\exp\left\{f_i(Z_i) - f_i(\tilde{Z}_i)\right\}\right] = \prod_{i=1}^n \left(\sum_{k=0}^{\infty} \frac{1}{k!} \mathbb{E}\left[\left(f_i(Z_i) - f_i(\tilde{Z}_i)\right)^k\right]\right) \\ & \stackrel{(c)}{=} \prod_{i=1}^n \left(\sum_{k=0}^{\infty} \frac{r^{2k}}{(2k)!} \mathbb{E}\left[\left(f_i(Z_i) - f_i(\tilde{Z}_i)\right)^{2k}\right]\right) \end{aligned} \quad (65)$$

In the above, step (a) follows from (64), step (b) uses the fact that  $\{\tilde{Z}_i\}$  is independent of  $\{Z_i\}$  and also they each contain independent elements, and step (c) holds since the odd order terms are zero. Now we bound the expectation term.

$$\begin{aligned} \mathbb{E}\left[\left(f_i(Z_i) - f_i(\tilde{Z}_i)\right)^{2k}\right] & \stackrel{(a)}{\leq} L_i^{2k} \mathbb{E}\left[\left((1 + \|Z_i\| + \|\tilde{Z}_i\|)\|Z_i - \tilde{Z}_i\|\right)^{2k}\right] \\ & \stackrel{(b)}{\leq} L^{2k} \mathbb{E}\left[\left(\|Z_i\| + \|\tilde{Z}_i\| + \|Z_i\|^2 + \|\tilde{Z}_i\|^2 + 2\|Z_i\|\|\tilde{Z}_i\|\right)^{2k}\right] \\ & \stackrel{(c)}{\leq} \frac{(5L)^{2k}}{5} \mathbb{E}\left[2\|Z_i\|^{2k} + 2\|Z_i\|^{4k} + 2^{2k}\|Z_i\|^{2k}\|\tilde{Z}_i\|^{2k}\right] \\ & \stackrel{(d)}{\leq} \frac{(5L)^{2k}}{5} \left(4(k)!(2d\nu)^k + 4(2k)!(2d\nu)^{2k} + 4(k!)^2(4d\nu)^{2k}\right), \end{aligned} \quad (66)$$

where step (a) holds due to the pseudo-Lipschitz property, step (b) uses the fact that  $L_i \leq L$  and  $\|Z_i - \tilde{Z}_i\| \leq \|Z_i\| + \|\tilde{Z}_i\|$ , step (c) uses Lemma B.5, and step (d) uses Lemma A.3 and another application of Lemma B.5. Plugging (66) into (65), we have

$$\begin{aligned} \mathbb{E} \left[ \exp \left\{ r \sum_{i=1}^n (f_i(Z_i) - \mathbb{E}[f_i(Z_i)]) \right\} \right] &\stackrel{(a)}{\leq} \frac{4}{5} \sum_{k=0}^{\infty} (25r^2 L^2)^k \left( (d\nu)^k + (4d^2\nu^2)^k + (8d^2\nu^2)^k \right) \\ &\leq \sum_{k=0}^{\infty} (25L^2)^k (d\nu + 12d^2\nu^2)^k = \frac{1}{1 - 25r^2 L^2 (d\nu + 12d^2\nu^2)} \stackrel{(b)}{\leq} \exp(50r^2 L^2 (d\nu + 12d^2\nu^2)), \end{aligned}$$

where step (a) uses the fact  $2^k(k!)^2 \leq (2k)!$  and step (b) uses  $(1-x)^{-1} \leq e^{2x}$  for  $x \in [0, 1/2]$ , which leads to the effective region for  $r$  to be  $r \leq (5\sqrt{2}L(d\nu + 12d^2\nu^2)^{1/2})^{-1}$ .  $\square$

## B Algebraic Inequalities

**Lemma B.1.** [9, Concentration of Sums] If random variables  $X_1, \dots, X_M$  satisfy  $P(|X_i| \geq \epsilon) \leq e^{-n\kappa_i\epsilon^2}$  for  $1 \leq i \leq M$ , then

$$P\left(\left|\sum_{i=1}^M X_i\right| \geq \epsilon\right) \leq \sum_{i=1}^M P\left(|X_i| \geq \frac{\epsilon}{M}\right) \leq M e^{-n(\min_i \kappa_i)\epsilon^2/M^2}.$$

**Lemma B.2.** [9, Concentration of Square Roots] Let  $c \neq 0$ . Then

$$\text{If } P(|X_n^2 - c^2| \geq \epsilon) \leq e^{-\kappa n\epsilon^2}, \text{ then } P(|X_n| - |c| \geq \epsilon) \leq e^{-\kappa n|c|^2\epsilon^2}.$$

**Lemma B.3.** [9, Concentration of Products] For random variables  $X, Y$  and non-zero constants  $c_X, c_Y$ , if

$$P(|X - c_X| \geq \epsilon) \leq K e^{-\kappa n\epsilon^2}, \quad \text{and} \quad P(|Y - c_Y| \geq \epsilon) \leq K e^{-\kappa n\epsilon^2},$$

then the probability  $P(|XY - c_X c_Y| \geq \epsilon)$  is bounded by

$$\begin{aligned} P\left(|X - c_X| \geq \min\left(\sqrt{\frac{\epsilon}{3}}, \frac{\epsilon}{3c_Y}\right)\right) + P\left(|Y - c_Y| \geq \min\left(\sqrt{\frac{\epsilon}{3}}, \frac{\epsilon}{3c_X}\right)\right) \\ \leq 2K \exp\left\{-\frac{\kappa n\epsilon^2}{9 \max(1, c_X^2, c_Y^2)}\right\}. \end{aligned}$$

**Lemma B.4.** [9, Concentration of Scalar Inverses] Assume  $c \neq 0$  and  $0 < \epsilon < 1$ .

$$\text{If } P(|X_n - c| \geq \epsilon) \leq e^{-\kappa n\epsilon^2}, \text{ then } P(|X_n^{-1} - c^{-1}| \geq \epsilon) \leq 2e^{-n\kappa\epsilon^2 c^2 \min\{c^2, 1\}/4}.$$

**Lemma B.5.** [9] For any scalars  $a_1, \dots, a_t$  and positive integer  $m$ , we have  $(|a_1| + \dots + |a_t|)^m \leq t^{m-1} \sum_{i=1}^t |a_i|^m$ . Consequently, for any vectors  $\underline{u}_1, \dots, \underline{u}_t \in \mathbb{R}^N$ ,  $\left\|\sum_{k=1}^t \underline{u}_k\right\|^2 \leq t \sum_{k=1}^t \|\underline{u}_k\|^2$ .

## C Pseudo-Lipschitz Function Lemmas

**Lemma C.1.** [9, Products of Lipschitz Functions are PL(2)] Let  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  be Lipschitz continuous. Then the product function  $h : \mathbb{R}^p \rightarrow \mathbb{R}$  defined as  $h(x) := f(x)g(x)$  is pseudo-Lipschitz of order 2.

**Lemma C.2.** [9] Let  $\phi : \mathbb{R}^{t+2} \rightarrow \mathbb{R}$  be  $PL(2)$ . Let  $(c_1, \dots, c_{t+1})$  be constants. The function  $\tilde{\phi} : \mathbb{R}^{t+1} \rightarrow \mathbb{R}$  defined as

$$\tilde{\phi}(v_1, \dots, v_t, w) = \mathbb{E}_Z \left[ \phi \left( v_1, \dots, v_t, \sum_{r=1}^t c_r v_r + c_{t+1} Z, w \right) \right] \quad (67)$$

where  $Z \sim \mathcal{N}(0, 1)$ , is then also  $PL(2)$ .

## References

- [1] Y. Ma, M. Y. Lu, and D. Baron, “Multiprocessor approximate message passing with column-wise partitioning,” in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Mar. 2017.
- [2] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, Aug. 2001.
- [3] Y. Zhou, U. Porwal, C. Zhang, H. Ngo, L. Nguyen, C. Ré, and V. Govindaraju, “Parallel feature selection inspired by group testing,” in *Neural Inf. Process. Syst. (NIPS)*, Dec. 2014, pp. 3554–3562.
- [4] X. Wang, D. Dunson, and C. Leng, “DECORrelated feature space partitioning for distributed sparse regression,” *Arxiv preprint arXiv:1602.02575*, Feb. 2016.
- [5] Z. Peng, M. Yan, and W. Yin, “Parallel and distributed sparse optimization,” in *Proc. IEEE 47th Asilomar Conf. Signals, Syst., and Comput.*, Nov. 2013, pp. 659–646.
- [6] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.
- [7] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing,” *Proc. Nat. Academy Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [8] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [9] C. Rush and R. Venkataramanan, “Finite sample analysis of approximate message passing,” *Proc. IEEE Int. Symp. Inf. Theory*, June 2015, full version: <https://arxiv.org/abs/1606.01800>.
- [10] P. Han, R. Niu, M. Ren, and Y. C. Eldar, “Distributed approximate message passing for sparse signal recovery,” in *Proc. IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, Dec. 2014, pp. 497–501.
- [11] P. Han, R. Niu, and Y. C. Eldar, “Modified distributed iterative hard thresholding,” in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 3766–3770.
- [12] P. Han, J. Zhu, R. Niu, and D. Baron, “Multi-processor approximate message passing using lossy compression,” in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016.

- [13] J. Zhu, A. Beirami, and D. Baron, “Performance trade-offs in multi-processor approximate message passing,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, July 2016.
- [14] J. Zhu and D. Baron, “Multi-processor approximate message passing with lossy compression,” *Arxiv preprint arXiv:1601.03790*, Jan. 2016.
- [15] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices,” *J. Stat. Mech. - Theory E.*, vol. 2012, no. 08, p. P08009, Aug. 2012.
- [16] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. OUP Oxford, 2013.